

# PATENT ABSTRACTS OF JAPAN

(11) Publication number : 11-282837  
(43) Date of publication of application : 15. 10. 1999

(51) Int. Cl. G06F 17/27

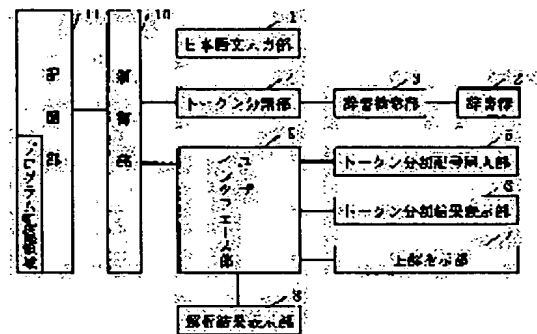
(21) Application number : 10-079010 (71) Applicant : MATSUSHITA ELECTRIC IND CO LTD  
(22) Date of filing : 26. 03. 1998 (72) Inventor : KINOSHITA HITOMI

(54) JAPANESE MORPHEME ANALYSIS DEVICE AND METHOD AND RECORDING MEDIUM

**(57) Abstract:**

**PROBLEM TO BE SOLVED:** To provide a Japanese morpheme analysis device which can obtain a correct interpretation via the instruction by a user even when plural interpretations are effective.

**SOLUTION:** This analysis device includes a Japanese sentence input part 1 which inputs Japanese sentences as character strings, a dictionary group 2 which stores Japanese words, the information on parts of speech of the Japanese words, the vocabulary information necessary for analysis of morphemes, etc., a dictionary retrieval part 3 which retrieves the group 2 with a Japanese word used as a key, a token dividing symbol insertion part 5 which indicates the token breaks of Japanese sentences, a token dividing part 4 which divides a character string inputted via the part 1 into tokens by referring to the dictionary information and the token breaks indicated via the part 5 and a control part 10 which totally controls the analysis device.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) ; 1998, 2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開平11-282837

(43) 公開日 平成11年(1999)10月15日

(51) Int.Cl.<sup>6</sup>

識別記号

F I

G 0 6 F 17/27

G 0 6 F 15/38

E

審査請求 未請求 請求項の数 5 O L (全 10 頁)

(21) 出願番号 特願平10-79010

(22) 出願日 平成10年(1998) 3 月26日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 木下 ひとみ

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

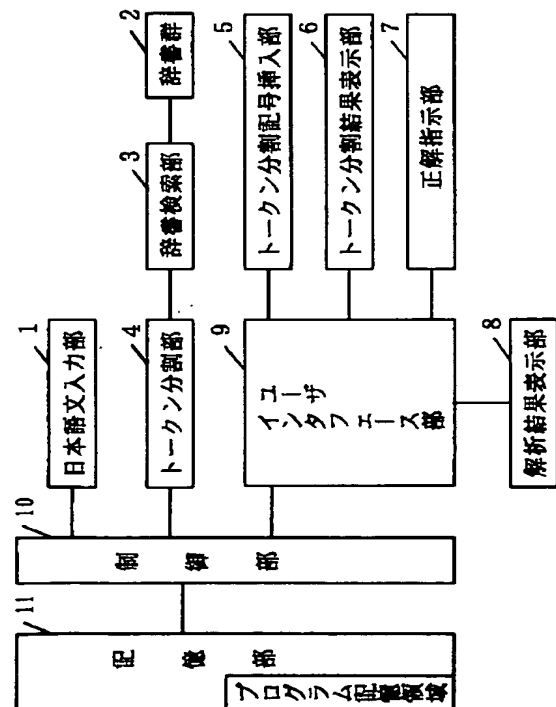
(74) 代理人 弁理士 滝本 智之 (外 1 名)

(54) 【発明の名称】 日本語形態素解析装置、日本語形態素解析方法および記録媒体

(57) 【要約】

【課題】 複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる日本語形態素解析装置を提供すること。

【解決手段】 日本語文を文字列として入力する日本語文入力部 1 と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群 2 と、日本語単語をキーとして辞書群 2 を検索する辞書検索部 3 と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部 5 と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部 5 で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部 4 と、全体を制御する制御部 10 とを有する。



## 【特許請求の範囲】

【請求項 1】日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして前記辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、前記日本語文入力部から入力された文字列を前記辞書情報と前記トークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することを特徴とする日本語形態素解析装置。

【請求項 2】日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして前記辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、前記日本語文入力部から入力された文字列を最初は前記辞書情報のみを用いてトークンに分割し、前記辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には前記曖昧性が生じた部分に対して前記辞書情報と前記トークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することを特徴とする日本語形態素解析装置。

【請求項 3】日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、前記入力ステップで入力された文字列を前記辞書情報と前記指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有することを特徴とする日本語形態素解析方法。

【請求項 4】日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、前記入力ステップで入力された文字列を最初は前記辞書情報のみを用いてトークンに分割し、前記辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に前記曖昧性が生じた部分に対して前記辞書情報と前記指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することを特徴とする日本語形態素解析方法。

【請求項 5】請求項 3 又は 4 に記載された日本語形態素解析方法を実行させるためのプログラムを記録した記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、文字列として入力した日本語文の形態情報を出力する日本語形態素解析装

置、日本語形態素解析方法およびその方法を実行させるためのプログラムを記録する記録媒体に関する。

## 【0002】

【従来の技術】ワープロのかな漢字変換や機械翻訳などの日本語処理においては、まず形態素解析を行う必要がある。形態素解析では普通、単語をキーとしてその語彙情報を記憶した辞書を検索しながら、文字列を形態素（意味を持つ最小の単位であり、この最小の単位を以下「トークン」と呼ぶ）に分割し（トークン分割）、このトークンに品詞、活用などの形態情報を付加する。形態素解析には、文節数最小法、左最長一致法、コスト最小法等の手法があり、これらの手法を用いて曖昧性を解消している。しかし、どの手法も完全ではなく、誤解釈を導くことがある。たとえば、「怪我しないように気を付ける」という日本語文の「怪我しないように」の部分をトークンに分割する場合を考えてみると、“怪我し／ないよう／に”という第 1 のトークン分割文、“怪我し／ない／よう／に”という第 2 のトークン分割文、“怪我／し／ないよう／に”という第 3 のトークン分割文、“怪我／し／ない／よう／に”という第 4 のトークン分割文などのように複数の解釈が存在することが分かる。尚、このトークン分割で参照した辞書には、サ行変格活用動詞「怪我する」、「する」、普通名詞「怪我」、「ないよう（内容）」、助動詞「ない」、形式名詞「よう」、格助詞「に」等が登録されているものとする。

【0003】例文の場合、正解は第 2 の分割文もしくは第 4 の分割文であるが、前述の 3 つの手法で評価してみると、文節数最小法では“怪我し／ないよう／に”という第 2 の分割文となり、左最長一致法では“怪我し／ないよう／に”という第 1 の分割文、接続コスト最小法ではコストの付け方によってどの分割文かが定まることとなり、誤解釈を出力することになる。また、これらのどの手法を採っても、経験則に依る所が大きく、多種多様な状況を表現し得る自然言語を処理する場合、誤解釈を導くことは避けられない。

## 【0004】

【発明が解決しようとする課題】このように、従来の日本語形態素解析方法では、前後の文脈情報を用いない限り正解を導き出すのは難しいという問題点を有している。また、文脈処理の技術は、実用化レベルに達していないのが現状である。

【0005】この日本語形態素解析装置、日本語形態素解析方法および記録媒体では、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができることが要求されている。

【0006】本発明は、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができ日本語形態素解析装置、および、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈が得られる日本語形態素解析方法、ならびに、その日本語形態

10

20

30

40

50

素解析方法が実現される記録媒体を提供することを目的とする。

#### 【0007】

【課題を解決するための手段】この課題を解決するために本発明の日本語形態素解析装置は、日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有する構成を備えている。

【0008】これにより、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる日本語形態素解析装置が得られる。

【0009】この課題を解決するための本発明の日本語形態素解析方法は、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有する構成を備えている。

【0010】これにより、複数の解釈を有する場合であっても、ユーザ指示により、正しい解釈が得られる日本語形態素解析方法が得られる。

【0011】この課題を解決するための記録媒体は、上記日本語形態素解析方法を実行させるためのプログラムを記録した構成を備えている。

【0012】これにより、上記日本語形態素解析方法が実現される記録媒体が得られる。

#### 【0013】

【発明の実施の形態】本発明の請求項1に記載の発明は、日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することとしたものであり、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性が解消されるという作用を有する。

【0014】請求項2に記載の発明は、日本語文を文字

列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には曖昧性が生じた部分に対して辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することとしたものであり、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担が軽減されるという作用を有する。

【0015】請求項3に記載の発明は、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してトークンに分割する分割ステップとを有することとしたものであり、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性が解消されるという作用を有する。

【0016】請求項4に記載の発明は、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、入力ステップで入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することとしたものであり、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担が軽減されるという作用を有する。

【0017】請求項5に記載の発明は、請求項3又は4に記載された日本語形態素解析方法を実行させるためのプログラムを記録することとしたものであり、記録したプログラムの実行により、請求項3又は4に記載された日本語形態素解析方法が実現されるという作用を有する。

【0018】以下、本発明の実施の形態について、図1～図8を参照しながら説明する。

（実施の形態1）図1は、本発明の実施の形態1による日本語形態素解析装置を示すブロック図である。

【0019】図1において、1は形態素解析処理対象の日本語文字列を入力する日本語文入力部、2は日本語単語

10

20

30

40

50

語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶している辞書群、3は日本語単語をキーとして辞書群2を検索する辞書検索部、4は辞書検索部3の検索結果およびユーザから指示されたトークンの切れ目の情報を参照して日本語文入力部1で入力された文字列をトークンに分割するトークン分割部、5は入力文のトークンの切れ目をユーザから指示してもらうトークン分割記号挿入部、6はトークン分割部4でトークンを分割した結果、曖昧性を含むものを表示し、ユーザからの正解入力の指示を仰ぐトークン分割結果表示部、7はトークン分割結果表示部6で表示したリストの中から、正解をユーザに指示してもらう正解指示部、8は形態素解析結果をユーザに示す解析結果表示部、9はトークン分割記号挿入部5、トークン分割結果表示部6、正解指示部7、解析結果表示部8を制御するユーザ・インタフェース部、10は日本語文入力部1、トークン分割部4、ユーザ・インタフェース部9を制御する制御部、11は

日本語文入力部1で入力された文、辞書検索部3の検索結果、トークン分割部4で分割されたトークンデータ、トークン分割記号挿入部5でユーザより指示されたトークンの切れ目の情報および日本語形態素解析結果を記憶する記憶部である。

【0020】ここで、トークン分割記号挿入部5におけるユーザからの指示は、トークンの切れ目を表す記号のユーザからの入力である（トークンの切れ目を表す記号として、半角の「`」を用いることにする）。また、トークン分割結果表示部6においては、曖昧性を含むものは、解釈可能なトークン分割結果を示すリストとして表示される。

【0021】（表1）に、本実施の形態で用いる辞書群2に記憶した辞書情報の一例を示す。

【0022】

【表1】

辞書見出し	品詞	活用型	活用形
さ	動詞	サ行変格	未然形
し	動詞	サ行変格	未然形
し	動詞	サ行変格	未然形
しろ	動詞	サ行変格	命令形
する	動詞	サ行変格	終止形
する	動詞	サ行変格	連体形
すれ	動詞	サ行変格	仮定形
せよ	動詞	サ行変格	命令形
な	助動詞	形容詞	語幹
ないよう	普通名詞	なし	なし
に	格助詞	なし	なし
よう	形式名詞	なし	なし
怪我	普通名詞	なし	なし
怪我	動詞	サ行変格	語幹

【0023】（表1）においては左から、「辞書見出し」「品詞」「活用型」「活用形」の順に示している。ここで、活用する語（動詞、形容詞、形容動詞、助動詞など）の場合は、辞書見出しとして、語幹（活用しない）部分のみを登録する。また、語幹が存在しないもの（「する」や「くる」など）は、その活用形を全て登録しておく。

【0024】図2は本発明の実施の形態1による日本語形態素解析装置の具体的構成を示すブロック図である。

【0025】図2において、21はキーボードやポインティング・デバイスなどの入力装置、22は陰極線管デ

ィスプレイ（CRT）や液晶ディスプレイ（LCD）などの表示装置、23は装置を制御する中央処理装置（CPU）、24はデータを一時的に記憶するランダム・アクセス・メモリ（RAM）、25はCPU23が実行するプログラムを格納するリード・オンリー・メモリ（ROM）、26は2次記憶装置、27は読取装置、28はCD-ROM等の記録媒体である。

【0026】ここで、図1の装置構成と図2の装置構成との関係を説明する。図1および図2において、日本語入力部1、トークン分割記号挿入部5、正解指示部7は入力装置21により実現され、トークン分割結果表示部

6、解析結果表示部8は表示装置22により、記憶部11はRAM24により実現される。また、辞書群2はRAM24、ROM25、2次記憶装置26のいずれかにより実現される。辞書検索部3、トークン分割部4、ユーザ・インタフェース部9、制御部10は、CPU23がRAM24およびROM25とデータのやり取りを行いながら、ROM25に記憶された各種のプログラムを実行することにより実現される。なお、本実施の形態では、CPU23がROM25に記憶されたプログラムを実行する形態をしているが、CPU23が実行するプログラムは、読取装置27を用い、CD-ROM（コンパクト・ディスク・リード・オンリー・メモリ）などの記録媒体28に記録されたプログラムを実行する形態であっても構わない。このように構成することにより、本実施の形態は汎用コンピュータなどにおいて容易に実現が可能となる。

【0027】以上のように構成された日本語形態素解析装置について、以下その動作を図3、図4、図6、図7に基づいて説明する。図3、図4は本実施の形態による日本語形態素解析装置の動作を示すフローチャートであり、図3は、入力文のトークンの切れ目をユーザから指示された後に、その情報を用いてトークン分割を行う処理の流れを示すものである。また、図6は本実施の形態における形態素解析結果を現すトークン・リストを示すリスト図であり、図7は本実施の形態における辞書検索結果を表すリストを示すリスト図である。なお、これらのフローチャートは、CPU23がROM25に記憶されたプログラムを実行する様子を示すものである。

【0028】図3においてまず、日本語文入力部1から形態素解析処理対象の日本語文（文字列）を入力する（S1）。ここでは、“怪我しないように気を付ける”という文（以下、「入力文a」という）が入力されたものとする。また、ここで入力される文は、全て全角文字とする（すなわち、半角カタカナやAsciiコードは含まない）。

【0029】次に、入力文aのトークンの切れ目をユーザが指示する（S2）。ここでは、半角の“ ”を用いてトークンの切れ目をユーザが入力する。従来例で前述したことから、入力文aの正しいトークンの切れ目は、“怪我し”ない”よう”に”気を付ける”。”の第1分割文、または“怪我”し”ない”よう”に”気を付ける”。”の第2分割文のいずれかである。しかし、ユーザがいつも正しくトークンを認識できるとは限らない。ここでは、以下のように、敢えて一部誤った指示を行ったものとする。

【0030】すなわち“怪我しないように”気を付ける”。”をユーザの指示とする。ステップ3～ステップ15では、ステップ2でのユーザの指示および辞書情報を用いて、入力文aのトークン分割を行う。

【0031】まずステップ3では、分割処理中の文字位

置を示す変数POSに初期値として1をセットする。次に、この変数POSと入力文aの文字数（文の長さ）を比較し（S4）、変数POSの方が大きければトークン分割結果を表示して（S15）、処理を終え、そうでなければステップ5以降の処理を行う。

【0032】ステップ5では、辞書検索キーを取得し、変数keyLenに検索キーの文字数をセットする。検索キーは、変数POSが示す文字位置から最初に見付かる”までの文字列となる。従って、“怪我しないように”が最初の検索キーで、keyLenは8となる。

【0033】次に、ステップ5で取得した検索キーで辞書群2を検索し（S6）、検索キーが辞書に存在したか否か（そのキーで検索成功か否か）をチェックし（S7）、存在したならばステップ13へ、存在しなければステップ8へ移る。

【0034】ステップ8では、検索キーを図4のフローチャートで示す辞書検索ルーチンへ渡し、辞書検索処理を行う。最初の検索キー“怪我しないように”は、（表1）に示した辞書に登録されていないため、ここでの処理対象となる。

【0035】次に、ステップ8の辞書検索の結果、トークン分割に曖昧性が生じたか否かチェックし（S9）、曖昧性があればステップ10へ、曖昧性がなければステップ12へ移る。

【0036】ステップ10では、解釈可能なトークン分割結果を全て表示する。ここでは、トークン分割結果の第1文として“怪我し／ないよう／に”、第2文として“怪我し／ない／よう／に”、第3文として“怪我／し／ないよう／に”、第4文として“怪我／し／ない／よう／に”という4種類の分割結果が表示される。

【0037】この4種類の中から、正しいものをユーザが指示する（S11）。次に、こうして得られた正しいトークン分割結果をトークン・リストに登録する（S12）。

【0038】次に、変数POSに今処理した検索キーの長さ（keyLen）を加え、ステップ4に戻る（S14）。次のサイクル（ステップ4～ステップ11）では、“ように”を処理することになる。つまり、ステップ5で、“ように”が検索キーとして取得される。こうして、変

数POSが入力文の長さを越えるまで、処理を続けた結果、得られるトークン分割結果（トークン・リスト）が図6に示したものである。

【0039】ステップ7から移行するステップ13では、検索結果（品詞や活用形等のトークン情報）をトークン・リストに登録する。トークン・リストの一例を図6に示す。トークン・リストの個々のセルが、分割したトークンを表し、隣り合うトークンは双方向のポインタで連結されている。そして、そのセル内には、対応するトークンの形態情報が格納されている。

【0040】次に、図4を用いて辞書検索処理の流れを

説明する。ここでの辞書検索では、入力文のある位置から始まる全てのトークン（辞書見出しと一致するもの）を切り出し、図 7 に示すようなリストとして出力する全解探索の手法を取る。実際は、この手法に加えて接続可否の検証を行ったりする必要があるが、本質的なものではないので、説明を省く。今、本ルーチンに、処理対象文字列として「怪我しないように」が渡されたものとする。

【0041】まず、処理中の文字位置を示す変数POSおよび検索キーの長さを示す変数keyLenに初期値として1をセットし（S21）、入力文字列の長さ（ここでは8文字）を示す変数lastPosに入力文字列の長さをセットする。次に、変数POSと変数lastPosを比較し、変数POSの方が大きければ処理を終え、そうでなければステップ23以降の処理を行う（S22）。

【0042】ステップ23では、変数keyLenと変数lastPosを比較し、変数keyLenの方が大きければステップ29へ、そうでなければステップ24へ移る。

【0043】ステップ24では、検索キーを取得する。検索キーは、変数POSが示す文字位置から変数keyLen分の長さの文字列である（最初の検索キーはPOS(=1)からkeyLen(=1)文字なので「怪」）。次に、ステップ24で取得した検索キーで辞書群2を検索する（S25）。次いで、検索キーが辞書に存在したか否か（そのキーで検索成功か否か）をチェックし（S26）、存在したならばステップ27へ、存在しなければステップ28へ移る。

【0044】ステップ27では、検索に成功した文字列およびその形態情報を図7のようにリスト化する。図7に示した図は、「怪我しないように」を辞書群2を用いてトークンに分割した結果を示したものである。それぞれのセルは辞書検索に成功した文字列であり、その検索結果およびトークン文字列の先頭と末尾の文字のオフセット（入力文の先頭からの文字位置）が形態情報として記憶される。更に、後接するトークンの候補へのポイントを可能な限り記憶する。例えば、いま変数POSが4で、「ない」が検索されたとする。それぞれのリストの最後のセルでそのトークン文字列の末尾の文字のオフセットがPOS-1の値と一致するセルを探し、そのセルの次に連結する。

【0045】ステップ28では、変数keyLenの値を1増やし、ステップ23へ戻る。ステップ23から移行するステップ29では、変数POSから始まる文字列が1つでも辞書検索に成功したか否かチェックし、検索成功が1つもなければステップ30へ、検索成功が1つでも存在すればステップ31へ移る。

【0046】ステップ30では、検索成功がない場合には変数POS番目の文字を未知語として図7のリストに登録する。次に、変数POSの値を1増やし（S31）、変数keyLenに1をセットし（S32）、ステップ22へ戻

る。

【0047】こうして最後まで処理した結果を図7にリストとして示す。以上のように本実施の形態によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群2を用い、まず日本語文を文字列として入力し、日本語文のトークンの切れ目をユーザが指示し、入力された日本語文の文字列を辞書情報とユーザが指示したトークンの切れ目とを参照してトークンに分割するようにしたことにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができる。

【0048】（実施の形態2）本発明の実施の形態による日本語形態素解析装置は図1、図2と同様の構成であるので、その説明は省略する。本実施の形態による日本語形態素解析装置が実施の形態1と異なるところは、実施の形態1ではステップ2に示すようにまずユーザがトークンの切れ目を指示するのに対し、まず装置がトークン分割を行い、その結果トークン分割に曖昧性が生じた場合にその曖昧性が生じた箇所をユーザに示し、その中からユーザが正解を選択するという点である。

【0049】以上のような本実施の形態について、図5を用いて説明する。図5は本実施の形態による日本語形態素解析装置の動作を示すフローチャートである。

【0050】図5においてまず、日本語文入力部1から、形態素解析処理対象の日本語文（文字列）を入力する（S41）。ここでも、実施の形態1と同様に、「怪我しないように気を付ける。」という文（以下、「入力文a」という）が入力されたものとする。また、ここで入力される文も、全て全角文字とする（すなわち、半角カタカナやAsciiコードは含まない）。次に、図4を用いて説明した辞書検索処理を行う（S42）。辞書検索の結果、トークン分割に曖昧性が生じたか否かチェックし、曖昧性があればステップ44へ、曖昧性がなければ、ステップ46へ移る（S43）。

【0051】ステップ44では、解釈可能なトークン分割結果を全て表示する。ここでは、トークン分割結果の第1文として「怪我し／ない／よう／に／気を付ける／。」、第2文として「怪我し／ない／よう／に／気を付ける／。」、第3文として「怪我／し／ないよう／に／気を付ける。」、第4文として「怪我／し／ない／よう／に／気を付ける／。」という4種類の分割結果が表示される。この中から、正しいものをユーザが指示する（S45）。そして、こうして得られたトークン分割結果および個々のトークンの形態情報を表示して、処理を終える（S46）。

【0052】以上のように本実施の形態によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等

の辞書情報を記憶する辞書群2を用い、日本語文を文字列として入力し、入力された日本語の文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割するようにしたことにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われ、ユーザの負担を軽減することができる。

#### 【0053】

【発明の効果】以上のように本発明の請求項1に記載の日本語形態素解析装置によれば、日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができるという有利な効果が得られる。

【0054】請求項2に記載の発明によれば、日本語文を文字列として入力する日本語文入力部と、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群と、日本語単語をキーとして辞書群を検索する辞書検索部と、日本語文のトークンの切れ目を指示するためのトークン分割記号挿入部と、日本語文入力部から入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には曖昧性が生じた部分に対して辞書情報とトークン分割記号挿入部で指示されたトークンの切れ目とを参照してトークンに分割するトークン分割部と、全体を制御する制御部とを有することにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われるので、ユーザの負担を軽減することができるという有利な効果が得られる。

【0055】本発明の請求項3に記載の日本語形態素解析方法によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、日本語文のトークンの切れ目を指示する指示ステップと、入力ステップで入力された文字列を辞書情報と指示ステップで指示したトークンの切れ目とを参照してト

クンに分割する分割ステップとを有することにより、入力された日本語文の文字列が辞書情報とユーザからの指示情報との両方に基づいてトークン分割されるので、日本語文のトークン分割の曖昧性を解消することができ、複数の解釈（トークン分割結果）を有する場合であっても、ユーザ指示により、正しい解釈を得ることができるという有利な効果が得られる。

【0056】請求項4に記載の発明によれば、日本語単語、その品詞情報、形態素解析に必要な語彙情報等の辞書情報を記憶する辞書群を用い、日本語文を文字列として入力する入力ステップと、入力ステップで入力された文字列を最初は辞書情報のみを用いてトークンに分割し、辞書情報のみを用いたトークン分割の結果として曖昧性が生じた場合には日本語文のトークンの切れ目を指示すると共に曖昧性が生じた部分に対して辞書情報と指示されたトークンの切れ目とを参照してトークンに分割する分割ステップとを有することにより、辞書情報のみを用いたトークン分割結果に曖昧性が生じた場合のみにユーザ指示が行われるので、ユーザの負担を軽減することができるという有利な効果が得られる。

【0057】本発明の請求項5に記載の記録媒体によれば、請求項3又は4に記載された日本語形態素解析方法を実行させるためのプログラムを記録することにより、記録したプログラムを実行すれば、請求項3又は4に記載された日本語形態素解析方法を実現することができるという有利な効果が得られる。

#### 【図面の簡単な説明】

【図1】本発明の実施の形態1による日本語形態素解析装置を示すブロック図

30 【図2】本発明の実施の形態1による日本語形態素解析装置の具体的構成を示すブロック図

【図3】本発明の実施の形態1による日本語形態素解析装置の動作を示すフローチャート

【図4】本発明の実施の形態1、2による日本語形態素解析装置の動作を示すフローチャート

【図5】本発明の実施の形態2による日本語形態素解析装置の動作を示すフローチャート

【図6】本発明の実施の形態1における形態素解析結果を現すトークン・リストを示すリスト図

40 【図7】本発明の実施の形態1、2における辞書検索結果を表すリストを示すリスト図

#### 【符号の説明】

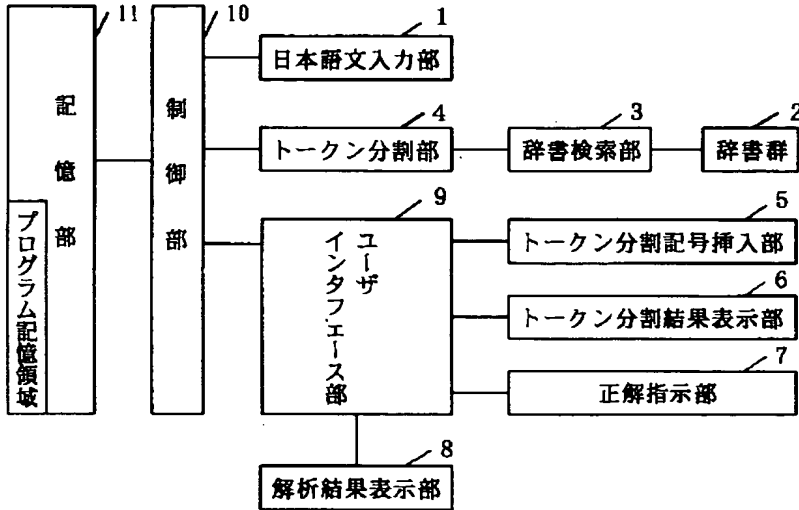
- 1 日本語文入力部
- 2 辞書群
- 3 辞書検索部
- 4 トークン分割部
- 5 トークン分割記号挿入部
- 6 トークン分割結果表示部
- 7 正解指示部
- 50 8 解析結果標示部



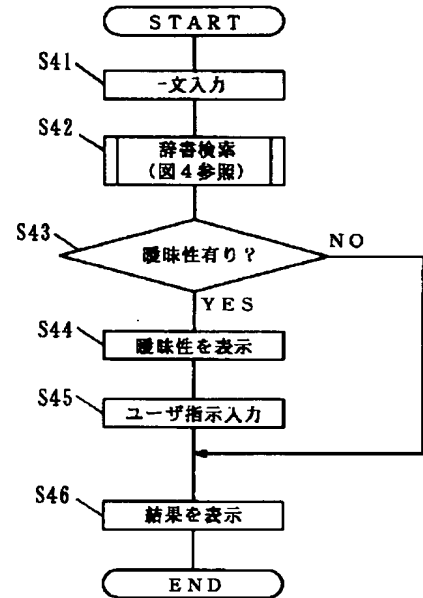
13  
9 ユーザ・インタフェース部  
10 制御部

11 記憶部

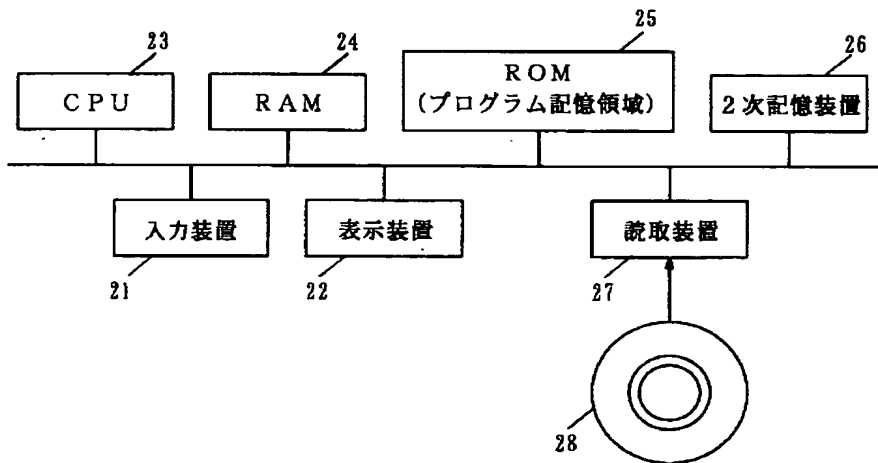
【図1】



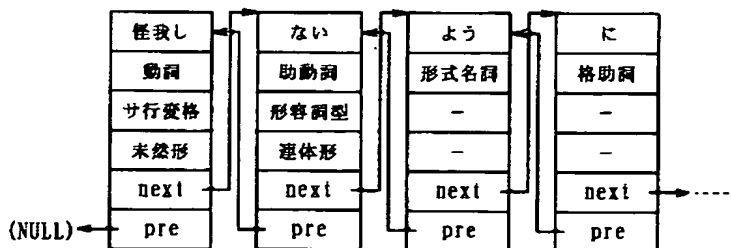
【図5】



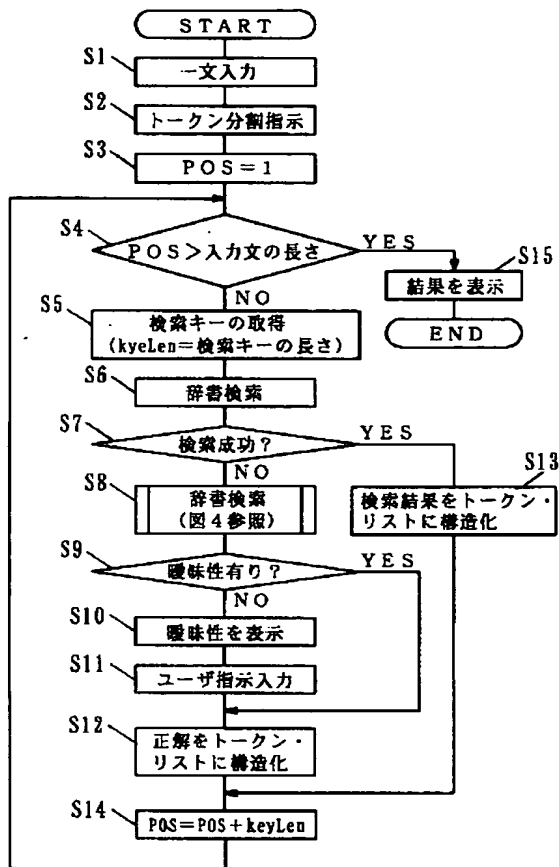
【図2】



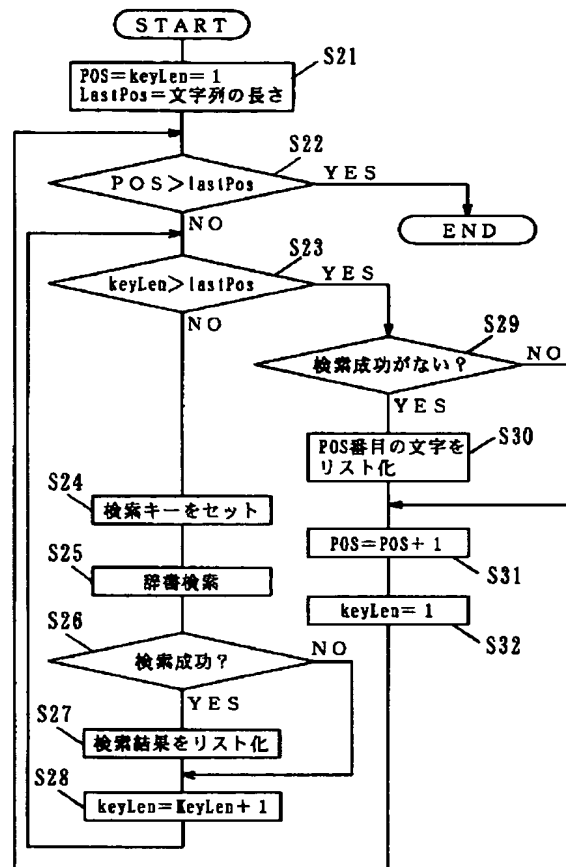
【図6】



【図3】



【図4】



【图7】

